AD-A096 649 DURHAM UNIV (ENGLAND) DEPT OF GEOGRAPHY F/G 9/2
STATISTICAL CHARACTERIZATION OF ALTITUDE MATRICES BY COMPUTER. --ETC(U)
1974 I S EVANS, I BAIN, H LEMONS DA-ERO-591-73-G-0040

UNCLASSIFIED NL

1 OF 1
AD A
096649

END
DATE
FILMED
4-81
DTIC

'STATISTICAL CHARACTERIZATION OF ALTITUDE MATRICES BY COMPUTER'

REPORT 2.

THE SECOND HALF-YEARLY PROGRESS REPORT ON GRANT DA-ERO-591-73-G0040

By IAN S. EVANS, M.A., M.S., PH.D., (Principal Investigator)

and IAIN BAIN, M.A. (Research Student)

Department of Geography, University of Durham, England.

To DR. H. LEMONS

Chief Scientist, European Research Office, U.S. Army.

1974

(1)

(6) STATISTICAL CHARACTERIZATION OF ALTITUDE MATRICES

BY COMPUTER.

REPORT 2. Calibration Tests on the Rayner-McCalden Program for 2-D Spectral Analysis.

THE SECOND HALF-YEARLY PROGRESS REPORT ON GRANT (15) DA-ERO-591-73-G0040
G-0040

(10)

By IAN S. EVANS, M.A., M.S., PH.D., (Principal Investigator)

and IAIN BAIN, M.A. (Research Student) (12) 37

[redacted] of Geography, University of Durham, England.

(9) Semi-annual progress rept. no. 2,

To DR. H. LEMONS

Chief Scientist, European Research Office, U.S. Army.

(11) 1974

INTRODUCTION

Progress has been made this year on two distinct aspects of the research project; the use of two-dimensional spectral analysis, and the calculation of vertical derivatives of a surface, starting in each case from an altitude matrix. The results are presented in modular form, with several reports on each aspect of the research. The first report deals with calibration of a spectral program, using artificial waveforms of known properties. Experience thus gained was used in re-running 2-D spectra for altitude matrices in the Digital Terrain Library, the subject of the second report. Thirdly, a summary of the value of 2-D spectra in general geomorphometry is offered. Two reports on vertical derivatives follow; the first presents preliminary results for the Digital Terrain Library, while the second summarises derivatives for different (photo-interpretation) terrain classes within a matrix area.

CONTENTS

CALIBRATION TESTS ON THE RAYNER-McCALDEN PROGRAM FOR 2-D SPECTRAL ANALYSIS

Initial results from using the 2-D spectral analysis program of Rayner and McCalden (1972) suggested a need for greater experience in selecting control parameters, to which the results are quite sensitive. The possibility of bias in the 2-D spectral domain (Evans and Bain, 1975) suggested the desirability of obtaining results for known inputs, such as cosine wave fluctuations. These trials firstly provide a test of program accuracy, and secondly suggest some approximate guidelines for control parameters. The outcome is to give the Rayner-McCalden program a remarkably clean bill of health.

The control parameters involved are G, the number of points smoothed ('tapered') at the edges of the data; (N-D), the number of zeros added to each row or column; and M^2, the number of spectral estimated required. D is the width of the data set (or the depth; we will deal here only with square matrices), which is generally given, while N is the width of the working array including the zero fill.

The degrees of freedom of an elementary spectral estimate are proportional to the square of (D-G)/N (Rayner, 1971, 1972). Increasing G or N decreases the degrees of freedom per elementary estimate, which is undesirable because it decreases the stability of such estimates. On the other hand, lack of edge smoothing and zero fill produces serious 'side lobes', spurious peaks in the spectral domain at harmonics of major peaks. Some compromise must therefore be made. Tukey (1961) suggested a wide range of possible values for G.

Similarly, increasing the number of elementary estimates grouped into a spectral band reduces the variance of spectral estimates, but increases band-width, i.e. decreases resolution (Jenkins, 1961). This 'smudging' may be reduced only at the expense of spectral stability. In the present experiments we test the effect of varying G and (N-D), but not M, which is held constant at 13. The effects of varying two properties of the generated surface, the wavelength and orientation of fluctuations, are also assessed. In each case, the other variables are held constant.

Hence experiments 1 and 2 are concerned with choice of parameters for running 2-D spectral analyses, while 3, 4 and 5 measure the accuracy of the program in locating oscillations of known properties.

All these experiments utilised 'dummy' input, for which the results could be predicted. A short program was written to create a 100 x 100 data matrix

containing a regular cosine fluctuation parallel to one axis, provision being made for alteration of wavelength and amplitude. By addition of a phase constant the program could rotate the cosine wave. It was also possible to create more complex surfaces by the addition of two or more wave components. Wavelength is expressed in data spacing units, i.e. a 100-unit wave goes through one cycle across the matrix.

1. Tapering (Edge smoothing)

With constant window width ((N-D) = 30), variance (≈50), wavelength and orientation of input, the extent of tapering was varied over the interval suggested by Tukey, i.e. from 4% to 25% of the data set length. Tukey's cosine bell function was used throughout.

Fig. 1 shows the percentage of variance remaining in the frequency domain, plotted against G expressed as a % of D. The graph shows an almost linear decline. As it stands, Fig. 1 represents the steeper portion of the graph of variance remaining against G. With a homogeneous data set the variance lost depends directly on the number of points affected by tapering. G rows and columns of the data set are tapered on each side, so each increase in G affects shorter rows and columns. Therefore each increase in G affects fewer additional points, and the rate of loss of variance should also decline: but this is not apparent for G less than 20%.

Somewhat different relationships would occur if variation in the data set were distributed unevenly. For example, if the edges of the data set exhibited relatively more variability than the central portions then the decline would be more rapid over lower values of G. Such a situation violates second-order stationarity, and would be difficult to remove by preliminary de-trending.

A measure of efficiency of the choice of parameters is given by the amount of variance allocated to the correct class in the frequency domain. Fig. 2 shows percentage variance in the correct class plotted as a function of G. A maximum of 93.35% occurs when G= 17% of D, when just under 30% of the original variance remains (with no tapering, 50% remains). However, it is unlikely that the best value of (G/D) will be constant for real, complex data sets. It is therefore necessary to try several different values, even though in a less controlled situation it may be very difficult to decide which result is 'best'.

2. Addition of Zeros

With constant tapering (G=10), variance (≈50), wavelength and orientation of

input, the number of zeros added to increase the data set length was altered. Not all values of N can be assessed, since the program requires that (N/2M) should be an odd integer greater than 1, where M is the number of estimates required. It was found that with N=100, i.e. with no zeros added, the variance in the correct class was 83.81%; with N=140 the variance was 91.16%. Variance remaining in the frequency domain was 62.58% and 31.58% respectively, but the 'loss of variance' due to addition of further zeros around the margins of a matrix is not as detrimental as the loss due to tapering. No information is lost: total variability remains the same, but since this is distributed over a larger number of points, variance is reduced. Hence the larger value of N is the more successful.

3. Wavelength

With constant window width ((N-D) = 30), tapering (G=10), variance (σ=50) and orientation the wavelength of the input was varied over a range that extended from 2.16 to 200 units. This demonstrated the program's ability to locate different frequencies precisely in the spectral domain, in both 1 and 2 dimensions. It was particularly useful to see the effects of fluctuations of wavelength greater than one quarter of the data set length, a condition analogous to that frequently met when the stationarity assumption is violated.

Initially a set of decreasing wavelengths was fed into the program; the results are given below.

Approximate wavelength (units)	Total variance in data set	% Total variance remaining after edge smoothing	% variance in 2-D spectral domain in correct cell
17.00	49.80	46.37	92.54
8.68	50.02	37.37	93.49
7.00	49.57	46.44	92.58
5.20	49.51	37.73	93.52
4.34	49.84	46.16	92.58
3.70	49.96	37.39	93.50
3.25	49.50	46.51	92.57
2.90	49.62	37.67	93.50
2.60	50.07	46.00	92.55
2.36	50.19	37.20	93.45
2.16	48.99	47.81	92.12

Differences in total variance are due to truncation of the cosine wave at different heights. That is, the wave always 'begins' at a maximum but its final value depends upon the relationship of the wavelength to the length of the data set.

Each wavelength chosen coincides with a class midpoint and it is clear from column four that the program works well in placing 92-94% of the variance in the spectrum into the correct class. At least a further 5% of the variance is concentrated in the eight cells immediately adjacent to the correct one. This sometimes leaves less than 2% of the variance distributed over the remaining 355 cells in the frequency domain specified here, so that 'leakage' from the correct wavelength and orientation does not appear serious. The shaded maps of the spectral domain, however, convey a more pessimistic impression. This is because their geometric class interval emphasises contrasts which involve very small proportions of variance (less than 0.01% per cell). In these simple examples the Rayner-McCalden program exhibits considerable accuracy. In each case, the 1-D averaged vector spectra peaks are quite sharp, and are located accurately.

A similar experiment was performed at higher resolution to see what happens to fluctuations whose wavelength falls near a class boundary.

The wavelength was altered in increments of 0.36 units over a range extending between two class mid-points in the frequency domain (6.5 and 8.07 units). The results are given in Fig. 3 where the two lines on the graph represent the differing percentages of variance in each class as the wavelength alters. It is notable that the decline away from each class mid-point is not linear. The values remain fairly constant until close to the class boundary, where rapid change occurs. Serious leakage of variance into adjacent cells occurs only for waves close to a cell boundary.

More interesting results are noted when an attempt is made to process waves too long for the program to resolve. The data set is 100 units long, so waves longer than 52 units should preferably be removed by detrending. Variance from these long-wave trends is placed in the 'zero frequency' class.

Wavelength (units)	Total variance in data set	Variance remaining after edge-smoothing	[illegible]
15	49.97	40.66	[illegible]
31	49.28	39.45	[illegible]
46	50.75	37.10	[illegible]
61	51.55	36.31	[illegible]
75	46.55	35.46	[illegible]
92	52.49	34.39	[illegible]
108	45.91	35.16	[illegible]
123	42.69	39.09	[illegible]
138	46.39	41.78	[illegible]
153	51.11	41.10	[illegible]
170	53.57	39.65	[illegible]
184	52.91	38.39	[illegible]
200	49.97	37.28	[illegible]

As before, differences in total variance are due to truncation; but in this case they are more extreme due to the longer wavelengths used.
The differences in the variance after edge-smoothing illustrate our earlier point about the uneven distribution of variance within a data set. It is noted that the greatest loss is for a 92-unit wave and the least is for 138 units. The data set is 100 units long and a maximum height occurs at one side of the matrix, hence a 92-unit wave reaches another maximum just before the other side of the matrix. Therefore much of the variance is concentrated at the edges of the data set and is affected by the smoothing function which reduces the greatest deviations. Although the cosine function starts at a maximum, when the period is 138 units it ends close to its mean value: edge-smoothing is then proportionately less important.
With regard to percentages of variance assigned to the zero frequency cell, a cut-off point should be observed at 52 units, which is the lower limit of the cell. With a wavelength of 31 units 5.43% is placed in this cell, at 46 units it is already 24.4%, while at 61 units only 70.45% is correctly placed. After a maximum of 91.3% at 75 units, the percentage correct decreases until at 200 units over 15% of the variance is leaked back into the frequency domain in a manner analogous to aliasing at the higher-frequency end of the spectrum. It seems, then, that resolution around this particular class boundary is not so good. Leakage into the zero frequency cell (e.g. for a 46-unit wave) is less serious than leakage out, which increases for wavelengths greater than [illegible] of the data width.

4. Orientation

With constant window width ((N-D) = 30), tapering (C=10) and variance of input, wave orientation was altered to test whether waves oblique to the axes can be resolved as accurately as those discussed so far. Waveforms were generated in the previous experiments by taking a cosine function as the first row, and repeating it the required number of times. To minimise additional computing, rotation is achieved here by adding a phase factor to the cosine fluctuation for each successive row. While this gives a rapid and efficient result, the wavelength is reduced as the angle of rotation increases, tending towards zero wavelength as ninety-degree rotation is approached. The adjusted wavelength can be calculated by multiplying the original value by the cosine of the angle of rotation. It is unfortunate that wavelength varies in this way; a slightly different algorithm can in fact avoid this without leading to any more computation. The results are given in Fig. 4. For each orientation a chart is given depicting the same subsection of the frequency domain. The index chart gives the wavelength of the midpoint of each class, while the series of charts gives percentage variance in the same classes. Again, the results are as expected, with serious leakage occuring only from waves close to class boundaries in the frequency domain. Leakage is, however, sufficiently strong to affect comparisons between cells with low proportions of total variance. Peaks in the 1-D averaged vector spectra are still sharp and precisely located, whatever the rotation.

5. Superimposed Waves

A further matrix was generated consisting of two superimposed cosine waves, with wavelengths of 20 units and 6.5 units and equal variance. These were aligned parallel to each other and to the east-west axis. In a second experiment a wavelength at 4.6 units was imposed on the 20-unit wave at an angle of 45^{o}. Results are given below:

WAVE COMPONENTS PARALLEL

Period (data units)	True Variance	Computed Variance	Correctly placed variance
20	50%	42.57%	42.57%
4.6	50%	50.14%	50.00%
			92.57%

WAVE COMPONENTS AT 45°

Period	True Variance	Computed Variance	Correctly placed variance
20	50%	46.71%	46.71%
4.6	50%	44.41%	44.41%
			91.12%

It can be seen from these results that the program is quite efficient at picking out the different wave components and allows less than 9% of the variance to be leaked into erroneous cells.

6. Conclusions

The Rayner-McCalden program is efficient at distinguishing wavelengths presented in the simple data sets used here. Leakage is confined mainly to the immediately adjacent cells and it can be seen that only when a wavelength comes close to the boundary does confusion occur, with variance being split between the classes on either side. This stresses the need to take account of the level of resolution required in selecting the number of estimates to represent the frequency domain.

Our previous suspicions (Evans and Bain, 1975 P. 5) concerning bias toward the principal axes of the frequency domain are withdrawn. The program works perfectly well with diagonal waves. Leakage does however occur preferentially parallel to the principal axes: this makes comparison of low peaks at different orientations very difficult.

The values of G, N and M can vary considerably. Higher values of G give better spectral accuracy but cause loss of variance. Higher values of N provide greater accuracy with countervailing disadvantage, but it must be remembered that the figure for total input variance is artificially lowered by inclusion of the extra zeros (i.e. mean values). The value of M (the number of estimates required in the frequency domain) depends on the resolution level desired and the number of degrees of freedom required per estimate. It seems likely that analysis of any particular data set might have to be repeated several times using different values of these parameters until a reliable result with good resolution is achieved.

Fig 1 REDUCTION IN VARIANCE AS A FUNCTION OF TAPER LENGTH

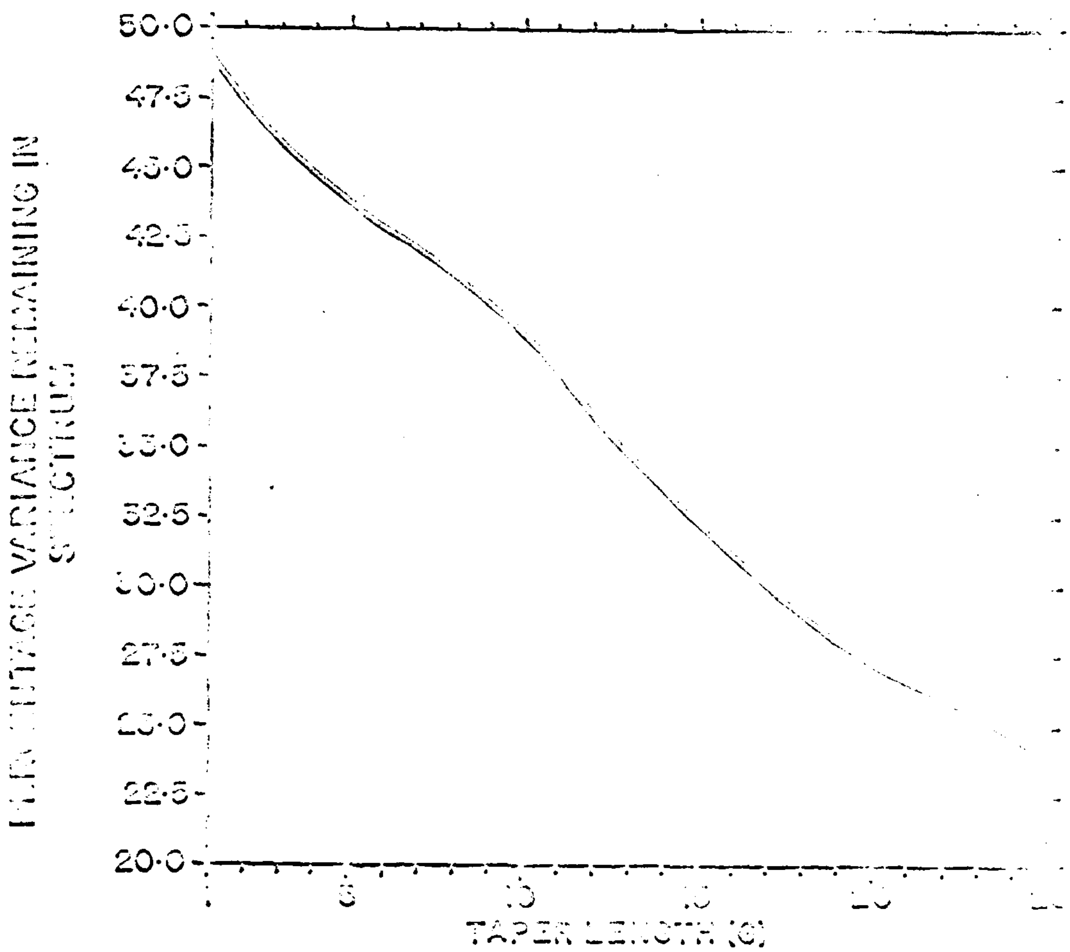

Fig 2 ACCURACY AS A FUNCTION OF TAPER LENGTH

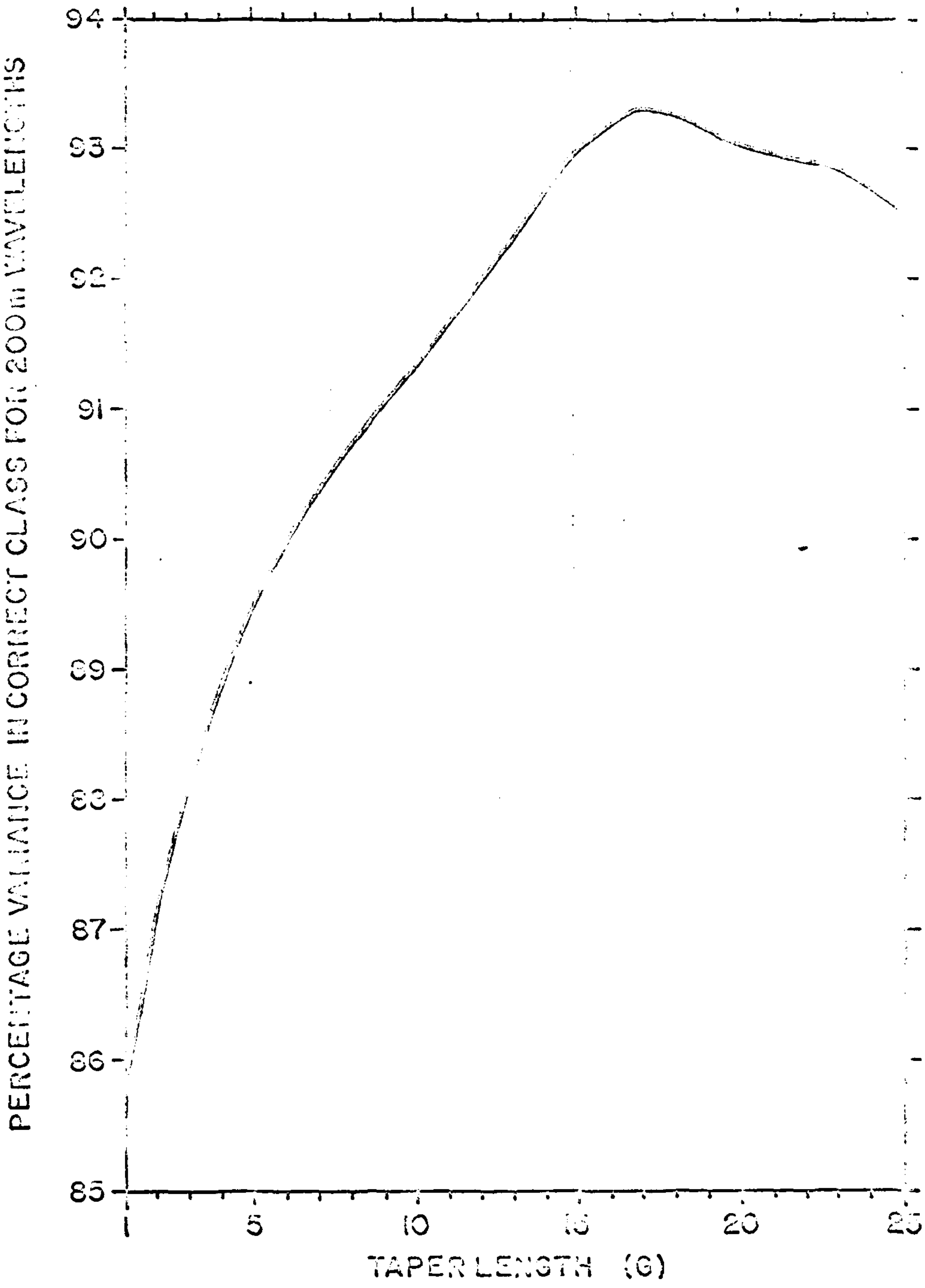

Fig 5 RESOLUTION AROUND A CLASS BOUNDARY

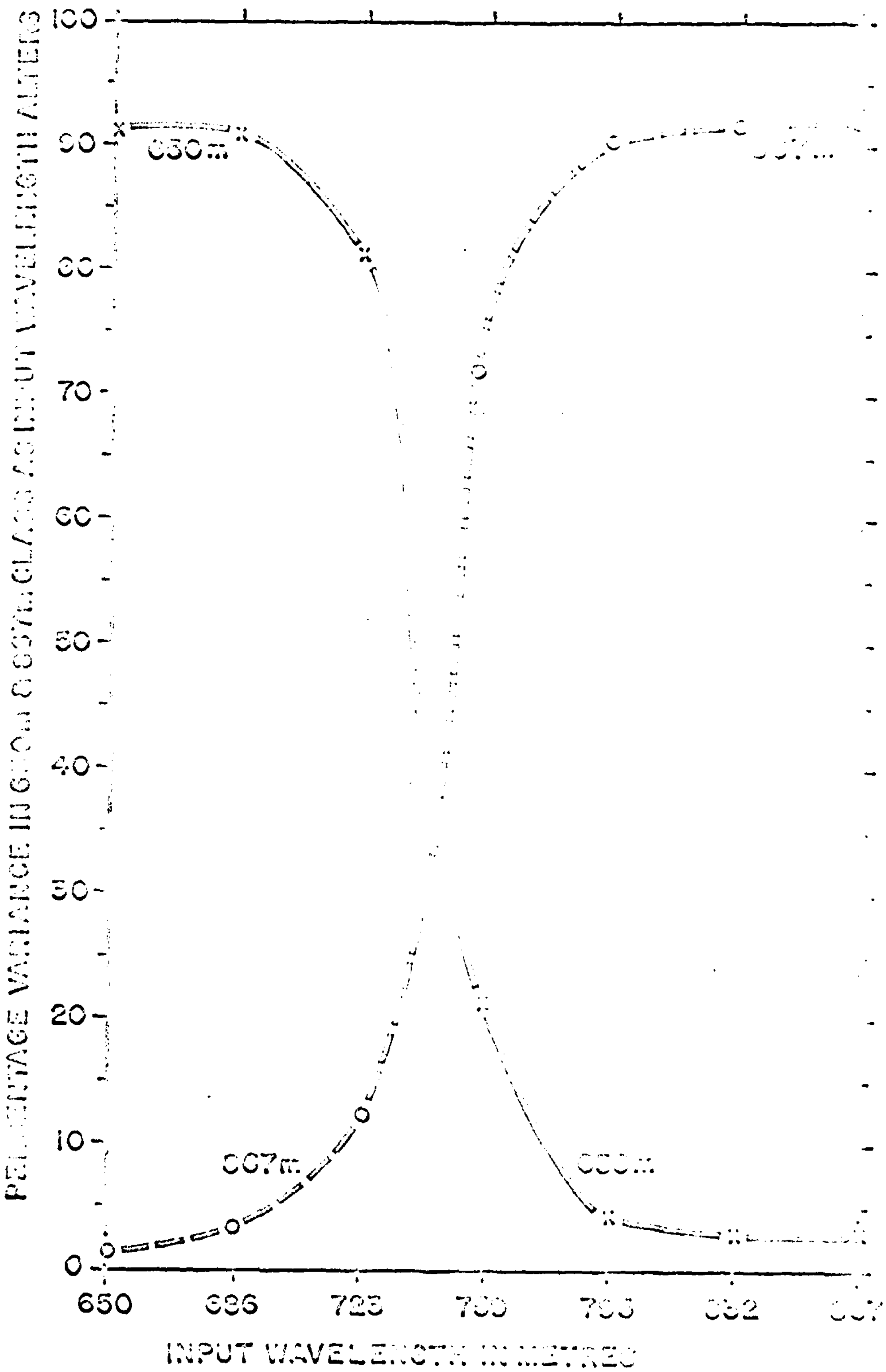

FIG 4

FREQUENCY DOMAIN MAPS

INDEX CHART. Numbers refer to wavelengths in units.

							class boundaries
0 FREQ	26.00	13.00	8.67	6.50	5.20	4.33	52.00
26.00	18.38	11.63	8.22	6.31	5.10	4.27	17.33
13.00	11.63	9.19	7.21	5.81	4.83	4.11	10.40
8.67	8.22	7.21	6.13	5.20	4.46	3.88	7.43
6.50	6.31	5.81	5.20	4.60	4.06	3.81	5.78
5.20	5.10	4.83	4.46	4.06	3.68	3.33	4.73

PERCENTAGE VARIANCE, 6·42 units at 5·2°

0·02	.	.	.	.	.	.
0·14	0·01	.	.	.	.	.
0·20	0·01	.	.	.	.	.
1·35	0·09	.	.	.	.	.
$\boxed{\times}$ 88·89	6.13	0·15	0·08	0·03	0·01	0·01
1·30	0·09	.	.	.	.	.

The boxed cross indicates the location of the input wave in the frequency domain.

PERCENTAGE VARIANCE, 6·40 units at 10·2°

·	0·01	·	·	·	·	·
·	0·09	·	·	·	·	·
0·01	0·20	·	·	·	·	·
0·04	1·35	0·01	·	·	·	·
3·15	91·43	0·88	0·17	0·01	0·01	·
0·05	1·35	0·01	·	·	·	·

PERCENTAGE VARIANCE, 6·27 units at 15·2°

·	0·01	·	·	·	·	·
·	0·09	·	·	·	·	·
·	0·20	0·01	·	·	·	·
0·02	1·35	0·04	·	·	·	·
1·39	91·[illegible]5 91·75	2·76	0·15	0·01	·	·
0·02	1·36	0·04	·	·	·	·

PERCENTAGE VARIANCE, 6·11 units at 19·0°

·	0·01	·	·	·	·	·
·	0·07	0·03	·	·	·	·
·	0·15	0·06	·	·	·	·
0·02	1·00	0·41	0·01	·	·	·
1·00	67·22	27·37	0·48	0·06	0·01	0·01
0·02	0·99	0·40	0·01	·	·	·

PERCENTAGE VARIANCE, 5·92 units at 24·3°

·	·	0·01	·	·	·	·
·	·	0·03	·	·	·	·
·	0·01	0·20	·	·	·	·
·	0·07	1·34	0·02	·	·	·
0·22	4·40	83·78 84·78	1·61	0·06	0·04	0·01
·	0·06	1·31	0·02	·	·	·

PERCENTAGE VARIANCE, 5·72 units at 20·4°

·	·	0·01	·	·	·	·
·	·	0·09	·	·	·	·
·	·	0·20	0·01	·	·	·
·	0·03	1·31	0·06	·	·	·
0·08	1·85	89·[illegible] [illegible]9·58	4·32	0·21	0·06	0·01
·	0·03	1·34	0·07	·	·	·

PERCENTAGE VARIANCE, 5·49 units at 32·3°

·	·	0·01	0·01	·	·	·
·	·	0·04	0·06	·	·	·
·	·	0·08	0·13	·	·	·
·	0·01	0·56	0·84	0·01	·	·
0·06	0·56	37·54	57·05	0·91	0·06	0·02
·	0·01	0·55	0·85	0·01	·	·

PERCENTAGE VARIANCE, 5·27 units at 35·0°

·	·	·	0·01	·	·	·
·	·	·	0·09	·	·	·
·	·	0·01	0·20	·	·	·
·	· ·	0·05	1·34	0·02	·	·
0·01	0·15	3·59	9[illegible]·79 90·79	1·55	0·15	0·01
·	·	0·05	1·34	0·02	·	·

PERCENTAGE VARIANCE, 5·04 units at 3[illegible]·1°

·	·	·	0·01	·	·	·
·	·	·	0·10	·	·	·
·	·	·	0·20	0·01	·	·
·	·	0·02	1·34	0·05	0·01	·
0·01	0·16	1·02	90·8[illegible] 90·82	3.67	0·50	0·02
·	·	0·01	1·35	0·05	0·01	·

[illegible]

With the knowledge gained from the tests with simple wave-form surfaces, the altitude matrices contained in the terrain library compiled by W.Tobler and the matrix for the Torridon area were re-processed. Generally 15% of the width was tapered at each edge of the matrix, and 15 rows or columns of zeroes were added at each edge. The spectral maps of four matrices showed clear side-lobes and were reprocessed either with broader tapering (Bray, Flaming Gorge), or with narrower tapering (Emerado, Delaware).

The effects of smoothing and the degree of non-stationarity of these data sets are presented below in Table 1, where the % total variance remaining in the frequency domain after edge smoothing is recorded along with the % of frequency domain variance which falls in the 'zero frequency' cell. For comparison the figures obtained in the first processing are also presented.

[illegible]

MATRIX	% TOTAL VARIANCE REMAINING IN FREQ. [illegible]		% FREQ. [illegible] IN 0 FREQ. CELL	
	[illegible]	1975	[illegible]	[illegible]
Emerado	13.36	45.10	75.61	[illegible]
Mammoth Cave	37.52	39.30	51.32	[illegible]
Flaming Gorge	21.14	46.41	76.09	[illegible]
Alma	25.32	45.44	25.30	[illegible]
M.Bonneville	35.[illegible]	41.26	62.53	73.01
Hillsboro	30.06	39.69	55.56	[illegible]
Ayer	36.33	45.46	78.15	[illegible]
Bray	29.53	54.42	[illegible]	[illegible]
Delaware	14.16	52.54	50.04	53.55
Camden	10.74	77.40	41.76	[illegible]
Maverick	27.65	40.12	[illegible]	[illegible]
Menan Buttes	22.45	47.57	[illegible]	[illegible]
Torridon	31.47	44.04	22.64	35.65

The figures for the % of total variance remaining in the frequency domain show how much variance is sacrificed in minimising side lobes. In

the case of Camden the reduction is sevenfold, since much of the variability of the terrain is situated near to the matrix edges. In the earlier processing the Camden matrix lost less than a quarter of its variance with C = 4. The greatest proportion of variance preserved in any matrix is just over a third (in the case of Ayer).

Another effect of the increased smoothing is to reduce the proportion of variance placed in the zero frequency cell, thus improving interpretability of the results. The reduction exceeds 10% of frequency domain variance for Torridon, Flaming Gorge and Menan Buttes. The only serious increase is for Bray (9.71%). Three areas are unaffected, and the other six have reductions between 2.91 and 8.19%. These figures indicate how much the smoothing operation affects the larger-scale forms in the matrices.

The incidence of the sidelobes (which the smoothing operations are designed to counteract) is related to the number of points in the matrix, the number of spectral estimates required, and the incidence of real peaks in the frequency domain. For the simulated cosine-wave matrices, variance 'leaked' from a peak in the frequency domain was concentrated into a cross centred on the peak; subsidiary peaks occurred at intervals determined by the size of the data set and the number of estimates required in the frequency domain. The pronounced bias toward the principal axes of the frequency domain observed previously (Evans and Lain, 1973), was due to the main spectral peak occurring in the central ('zero frequency') cell. Table II shows the bias of the frequency domains of the matrices, as measured by a ratio of variance on the principal axes to variance on the diagonals. ('Bias' = error consistently in one direction). The early results are shown for comparison.

In ten of the thirteen matrices there is a reduction in the degree of bias, which must be attributed to the increased edge smoothing. That bias remains is due to two factors; edge smoothing may not be effective enough , and the forms in the matrices may really vary with orientation. Increased smoothing has reduced the largest biases, by 50% for Emerado and

Delaware, and by 60% for Bray. Elsewhere, however, the effect of edge-smoothing is not great. It appears that considerable 'leakage' is unavoidable.

TABLE II

MATRIX	AVERAGE HORIZONTAL AXIS	AVERAGE VERTICAL AXIS	AVE. RIGHT DIAG.	AVE. LEFT DIAG.	BIAS (1974)	BIAS (1978)
Emerado	0.945	0.793	0.004	0.085	27.00	55.00
Mammoth Cave	1.369	1.013	0.194	0.344	4.45	6.53
Flaming Gorge	0.872	0.748	0.119	0.042	10.12	9.39
Alma	2.000	0.960	0.443	0.787	2.41	2.31
Mt. Bonneville	1.636	0.635	0.134	0.139	5.38	6.02
Hillsboro	1.360	0.574	0.266	0.494	2.55	2.74
Ayer	0.647	0.466	0.438	0.046	2.30	2.76
Bray	0.221	0.426	0.062	0.072	4.81	12.17
Delaware	0.634	2.303	0.267	0.170	7.73	15.31
Camden	0.956	1.493	1.331	0.111	1.64	1.11
Maverick	1.250	1.037	0.125	0.080	11.11	13.14
Menan Buttes	1.143	1.351	0.147	0.176	7.72	8.98
Torridon	2.367	2.170	0.330	0.653	4.62	3.22

The average values for the axes go some way to providing a description of the general character of the frequency domains, providing one takes into account the bias involved; possibly it is best in this case to compare the diagonals and principal axes separately. Thus we see that in the case of Emerado, north-south lineation is slightly stronger than east-west, while NW-SE lineation is much stronger than NE-SW.

If we assume that concentration of variance on the principal axes is not a real feature of the data, but is due solely to leakage, the bias might be overcome by dividing values on the principal axes by this measure of average bias. This seems reasonable here since many of the examples violate the stationarity requirement to a considerable degree, and their frequency domains are dominated by peaks in the central cell. If dominant peaks are in other cells, as for the stationary cosine waveforms analysed above,

bias would be calculated with reference to axes passing through the variance peak.

The problem of using these figures is that they are averages and they take no account of scale distribution of variance along the axes of the frequency domain. Moreover, they treat only four orientations from a theoretically infinite set. The choice of these four orientations is conditioned by the fact that they cut through the mid-points of cells. Profiles along other orientations would involve interpolation, based on debatable assumptions.

Assessment of the individual spectral maps is made difficult by the dominance of the central cell, but in general it is found that the longer wavelengths predominate. The peaks with which spectral analysis is often concerned are absent from most of these matrices. Only Camden and Emerado show concentration of variance on diagonals. Torridon has a peak at about 2600m, indicating a predominance of waves in the N-S direction: Menan Buttes has a small peak at about 880 feet.

Elsewhere differences are more subtle. In the case of Alma the longer wavelengths are pushed slightly towards the left side of the frequency domain, while the opposite occurs in the case of Ayer.

The problems of obtaining some measure that will describe frequency domains concisely and enable them to be compared with each other must await satisfactory detrending of the origin l data. At the moment the spectra show a reasonable improvement over the initial results, indicating the choice of better control parameters.

The difficulty, induced by this leakage of variance, in interpreting two-dimensional spectra encourages a closer inspection of one-dimensional 'collapsed vector spectra', in which the directional element is averaged out and variance is plotted against frequency. For the cosine waveforms, variance is strongly concentrated around the known true frequency. Peaks are about 10 estimates (out of 65) wide at a variance .01 that of the peak. Sidelobe peak variances are less than .004 of true peak height. This shows the accuracy of the Bowman-McCalden program in placing strong peaks within

the resolvable range of frequencies.

Collapsed vector spectra for the terrain library matrices all show a linear decline of log (variance) with increasing frequency, from a peak in the second spectral estimate: the first estimate, extending to zero frequency, now has less variance than the second. The spectra can therefore be described by (i) their height, i.e. total variance or standard deviation, and (ii) the logarithmic slope of the decline. A high rate indicates greater concentration at low frequencies (long wavelengths). Because of arbitrary alterations in spectral processing, it is best to take original variance for (i), rather than variance in the spectrum.

Because of the low first estimate, and a curve-over around the peak, a regression is not fitted to the whole set of spectral estimates. Rough estimates of the rate of decline are obtained by taking logarithms of ratios of variance in the 4th, 30th, 31st and 40th estimates; the replication provides some idea of the variability in these ratios(Table III). Standard deviation of the original altitude data is a measure of overall surface roughness. Since larger areas tend to be more variable than smaller, the varying areas covered by these matrices hinder comparison. The major contrasts, however, are of a larger order than the appropriate correction; hence Flaming Gorge and Torridon can be considered rougher than Menan Buttes and the Mammoth Cave area.

The ratios from the spectra show some variation between runs with different control parameters, but the results are reasonably consistent. Hence Flaming Gorge, Ayer and Bray have dominantly longwave variability, compared with more important shortwave roughness in Alma and Hillsboro. Although these ratios tend to be higher for rougher surfaces, they are sufficiently independent to separate areas of similar overall roughness.

The collapsed vector spectra therefore provide one useful terrain descriptor, the logarithmic gradient of the plot of variance against frequency. Perhaps the best expression of this would be to fit a regression to spectral estimates between say the 3rd and 40th.

TABLE III

	G	M	N	$\log\left(\frac{4th}{30th}\right)$	$\log\left(\frac{4th}{31st}\right)$	$\log\left(\frac{4th}{40th}\right)$	Standard deviation (feet)	Grid mesh (feet)	Matrix size
Emerado	10	11	110	4.28	4.39	4.67	151	651	80
	15	11	110	4.14	4.27	4.56			
	20	11	110	3.90	4.03	4.34			
	10	13	130	4.08	4.33	4.47			
Mammoth Cave	11	13	130	2.83	2.74	3.13	38	260	100
Flaming Gorge	11	13	130	4.44	4.63	4.75	606	100	100
	15	13	130	4.45	4.60	4.74			
	20	13	130	4.36	4.48	4.63			
Alma	10	11	110	1.85	1.91	2.32	151	521	80
Mt. Bonneville	11	13	130	3.97	4.06	4.53	[illegible]	130	100
Hillsboro	10	11	110	2.42	2.45	2.73	44	200	80
Ayer	11	13	130	4.14	4.25	4.66	66	100	100
Bray	15	13	130	4.31	4.52	4.61	410	260	100
	20	13	130	4.34	4.60	4.65			
Delaware	10	11	110	3.49	3.53	3.70	104	200	80
	15	11	110	3.43	3.47	3.75			
	20	11	110	3.42	3.46	3.71			
	10	13	130	3.08	3.10	4.47			
Camden	10	13	130	3.13	3.00	2.99	7	250	80
Maverick	10	11	110	3.32	3.35	3.64	135	200	80
Menan Buttes	10	11	110	2.99	3.05	3.38	21	200	80
Torridon	10	13	130	3.93	4.02	4.09	504	328	100

A SUMMARY OF THE VALUE OF 2-D SPECTRA IN GENERAL [illegible]

(i) The main limitation on the empirical use of spectral analysis for comparing different terrains is the subjectivity of the results. These are critically dependent upon parameters such as degree of tapering, tapering function, window size and degree of smoothing or detrending. Rules of thumb are available for selecting parameters, but the final choice depends on a trade-off of variance loss against sidelobe suppression. Different decisions will be made for different spatial series, and by different researchers for the same series.

(ii) A spectral analysis program cannot be used as a 'package' by someone with little experience of its use. It is necessary to gain familiarity with the steps involved, and with the effects of varying the control parameters. These comments apply to one-dimensional analysis and even more strongly to two-dimensional.

(iii) The trend of increasing variance with increasing wavelength dominates all altitude spectra produced so far. Differences in the gradient of this trend, and in azimuthal variation, are much less evident and are sensitive to the control parameters. 'Leakage' parallel to the principal axes of the 2-D spectral domain produces a bias which is difficult to correct, and which interferes with the measurement of azimuthal variation. Although the absolute amount of variance leaking is small, it is large relative to that in neighbouring cells. In fact, this effect is so serious that separate 1-D analysis of differently-oriented profiles may be preferable to 2-D analysis of a square altitude matrix.

(iv) To give a significant spectral peak, a waveform must repeat at least several times in the data series. Many series contain two or three waves which appear strong visually, but are unlikely to be important in the spectrum.

(v) Manipulations in the spectral domain, to calculate for example gradient and convexity, are inadvisable because the spectrum has lost a considerable amount of information through tapering and grouping. Hence

transformations back into the spatial domain are never exact. The theoretical arguments in favour of spectral manipulation are overwhelmed by practical problems of producing good spectra from finite and probably non-stationary records. Much more precise results are obtained by calculating derivatives in the spatial domain.

(vi) Compared with other statistics describing profiles or surfaces, spectral estimates have the advantage of varying independently of each other. Nevertheless, the individual spectral estimates are not useful terrain descriptors. Gradient, aspect, profile and plan convexity are directly useful descriptors, relevant to geomorphologic theories and modelling. Moreover, the relations between these surface derivatives, far from being a disadvantage, form an important part of terrain description and explanation revealing, for example, the way gradient varies with aspect.

(vii) Although many geomorphic landscapes appear to have a periodic valley spacing, spectra are complicated by (a) valley curvature, blurring directional components, and (b) valley confluence which necessarily varies ridge height and spacing, blurring spatial frequencies. Spectra will vary not only with geology but also with position in the erosional system, so that it is almost impossible to meet stationarity requirements even after trends in the mean have been removed.

(viii) Harmonic (Fourier) analysis and auto-regressive modelling of surfaces are open to similar objections. The only realistic type of modelling would involve progressive spatial variation in the parameters of an autoregressive model. Even so, it may not be possible to model a fluvially-eroded surface as a continuous surface (Boehm, 1967), without taking into account the linear distribution of the drainage net.

PRELIMINARY CALCULATIONS OF VERTICAL DERIVATIVES

Two mathematical properties of surfaces which are of immediate concern to the geomorphologist are their first and second derivatives, slope and convexity (curvature) respectively. Here we shall consider the first vertical derivative, gradient. Calculation of the gradient at a point involves differentiating altitude with respect to two orthogonal horizontal axes, and taking the vector sum of the two.

$$(1) \quad z' = \sqrt{\left(\frac{dz}{dx}\right)^2 + \left(\frac{dz}{dy}\right)^2}$$

There are two main procedures for obtaining derivatives, given that the surface is presented in the form of an altitude matrix. The first and more complex method is to fit a polynomial to a small neighbourhood of altitudes, centred on the point for which one wishes to estimate the derivatives. It is necessary to keep the number of points to a minimum to avoid excessive averaging of gradient, losing spatial detail. To obtain first and second derivatives a quadratic polynomial is required, and a neighbourhood of nine points arranged in a 3 x 3 sub-matrix is sufficient to calculate this.

The second and more direct method is to employ the calculus of finite differences. This involves calculating the difference between points adjacent to that for which the derivative is to be obtained, in the x and y directions, and then calculating the vector sum of these. The operation is expressed as

$$(2) \quad z'_{ij} = \sqrt{(z_{i,j+1} - z_{i,j-1})^2 + (z_{i+1,j} - z_{i-1,j})^2}$$

This averaged difference can be expressed as a tangent by dividing by twice the sampling interval of the original matrix. The second vertical derivative is obtained by performing operation (2) on the matrix of first derivatives.

Both methods have distinct advantages and disadvantages. The finite difference technique has the advantage of a very simple algorithm.

Calculating a quadratic polynomial is more efficient in retaining resolution as only 9 points are required to get the second derivative. Finite differences involve 15 points, meaning that an estimate from this method is an average for a much wider area.

For our initial experiments with spatial derivatives the finite difference technique was chosen for computational simplicity. W. Tobler used this method in a study of land classification (1969) and has published a FORTRAN program for the calculation of spatial derivatives (Tobler, 1970). A program was written to calculate derivatives and parametric statistical descriptors (mean, range, standard deviation, skewness and kurtosis) of their frequency distributions. The following table gives these statistics of altitude and of gradient (first vertical derivative) for matrices in the digital terrain library, excluding Mount Bonneville. The expectations of both skewness and kurtosis are zero for a normal distribution.

	ALTITUDE (feet)					GRADIENT (degrees)				
	RANGE	MEAN	ST.DEV.	SKEW.	KURTOSIS	RANGE	MEAN	ST.DEV.	SKEW	KURTOSIS
Emerado	245	913	43	+0.72	+0.01	8	1.45	1.35	+0.66	+3.19
Mammoth Cave	420	637	38	+1.27	+6.38	44	5.36	4.45	+1.30	+7.73
Flaming Gorge	1880	6607	606	+0.38	-1.15	67	26.23	14.68	-0.27	+2.19
Alma	600	1016	151	-0.16	-1.14	71	32.36	15.32	-0.12	+2.08
Hillsboro	220	866	44	-0.23	-0.58	32	9.57	4.73	+0.48	+3.18
Ayer	390	336	88	+1.05	+0.52	32	5.51	4.28	+1.21	+5.35
Bray	2440	5875	411	-0.40	+0.42	54	17.52	12.54	+0.47	+2.45
Delaware	635	1093	104	-1.02	+2.00	51	10.02	8.10	+1.19	+4.02
Camden	120	699	7	+5.02	+63.78	29	0.50	1.59	+6.52	+70.64
Maverick	680	6479	135	+0.31	- 0.40	53	13.22	8.64	+1.24	+4.72
Menan Buttes	105	4840	21	+1.42	+1.25	12	2.35	1.90	+1.04	+4.37
Torridon	1076	184	161	+1.39	+1.68					

The correlation between mean gradient and standard deviation of altitude is significant at the 95% confidence level ($r_s = +.87$, $n = 11$). There is a very strong correlation ($r_s = +.99$) between mean and standard deviation of gradient, and moderate (significant) correlations between standard deviation

and skewness both for gradient (r_s = +.59) and for altitude (r_s = +.67). High skewness is necessarily accompanied by high kurtosis, but these two statistics otherwise vary almost independently. Comparison is hindered by the varying matrix resolutions, but the information tabulated here does seem to provide a reasonable discrimination between the areas.

RELATION OF FIELD INFORMATION TO THE TORRIDON ALTITUDE MATRIX

Matrix overlaying

Relationships between process and form are a central theme in geomorphology. Altitude matrices of sufficient resolution provide specific information on form, but relationships that can be inferred with process are at best very general. Links between process and form can, however, be established by comparison with other geographic matrices.

In an altitude matrix, altitudes at regular intervals are stored with geographic locations implied by the row and column subscripts. In exactly the same way, other geomorphic distributions may be point-sampled on a square grid and stored in matrix form. If for a given area, altitude and other variables are sampled at the same points, then the matrices may be overlaid and compared in a computer version of the traditional geographic visual technique of map comparison.

Such a study has been made and is described below. Its basis is an altitude matrix which had been derived for part of the Torridon mountain area in North-west Scotland. The area of the matrix is 100 km^2 and heights were sampled at intervals of 100 m. Terrain information is required, then, at these 10,000 points : clearly the information must be highly generalised.

A 'terrain matrix' for the same area was derived from a map based on air photographic interpretation by I. Bain. The technique was comparable to that used by Goodier and Grimes (1970) in North Wales, but focussed on terrain types rather than vegetation. The types were distinguished in terms of photographic texture and tone, although the latter is unduly influenced by vegetation. A further difference is that the air photos which could be obtained here were on a 1:27,000 scale, whereas Goodier and Grimes used 1:10,000 air photos. Despite the vegetation 'noise factor', the resulting map relates to surface textures and forms at a scale much finer than those captured by the 100 m mesh altitude matrix. These features may perhaps be used as morphologic surrogates for process.

Generation of a terrain matrix

The mapping of the area was aided by the existence, for

the neighbouring Beinn Eighe Nature Reserve, of terrain maps created by Sargeant and collaborators in the Nature Conservancy. 1:10,000 photo coverage permitted them to draw detailed maps of terrain types. Although only 13 basic terrain types were distinguished, these were combined in different ways, resulting in a much larger number of areal classes; this makes further processing difficult.

To simplify coding for computer operations, the number of map classes was kept as small as possible. The generalisation involved seemed appropriate in relation to the smaller scale (1:27,000) of air photos available for the present study.

Careful scrutiny of the photographs coupled with some previous field knowledge of the area suggested recognition of ten terrain classes. The edges of the original altitude matrix could be pinpointed easily on the photographs. Using a Zeiss Sketchmaster, the terrain class boundaries were mapped on a base consisting of 1:10,000 and 1:10,560 photogrammetric maps. This terrain map was then sampled at 100 m intervals to give the 'terrain matrix'.

I. Bain spent two weeks in the field in September 1973 to check the validity of the terrain class boundaries defined on the photographs. This period was spent largely in the south of the area, where terrain of each class was accessible. It was found that the mapping of the class boundaries was quite accurate and only minor adjustments were required.

Terrain types

The area covered by this matrix has a uniform lithology consisting predominantly of Torridonian Sandstone, a hard coarse-grained arkosic grit with near-horizontal bedding. Tectonic activity, presumably associated with major thrust planes to the east of the area, has produced numerous small faults which combine with the pronounced bedding planes to facilitate erosion by block removal. During the Late Pleistocene, the area was extensively glaciated and during the last readvance period (Zone III) it probably supported a small ice cap. The area now consists of an upland plateau surrounded on three sides by major glaciated valleys. From this plateau rise several steep-sided peaks and ridges separated by well-defined troughs.

Vegetation is sparse throughout the area.

Brief descriptions of the terrain classes are given below.

0. Water

Areas were defined as 'water' only when points adjacent to that being considered were also located on water. Therefore this category refers to bodies of standing water generally larger than 100 m across.

1. Bare Rock

Eroded areas of bare rock are easily distinguished by their light tone on these aerial photographs. Where the gradient is slight, the rock outcrops in broad, fairly smooth pavements. With steeper gradients, erosion of alternating tough and weak lithologies creates a rapid alternation between sheer cliff faces and screes. Although they contain much detritus these areas have generally been designated bare rock.

2. Summit terrain

Summit terrain is distinguished by a light photographic tone. It consists of wide areas overgrown by moss (*Rhacomitrium spp.*), which give way to grass with decreasing altitude or increasing shelter. In places the moss heath gives way to bare areas of coarse sand and a 'clitter' of highly angular, comminuted debris. Finally, pavement areas of rocks weathering in situ and exhibiting signs of spalling exist, and in places there is a limited tor development. In one location (Beinn Alligin), terracette development was observed on a very gentle slope. Elsewhere in the Torridon area Sargeant (personal communication) has observed periglacial forms, and on Ruadh Stac Mor of Beinn Eighe these are being studied by A. Loades of the North London Polytechnic.

3. Steep scree

This class is recognised partly by the appearance of bare unvegetated scree slopes or by the growth of *Calluna* on steep screes. Although growth of heather requires a certain degree of stability, soil development is minimal. Generally the heather clings to a few centimetres of peaty soil, often on individual boulders. Where the scree consists of quartzite (Beinn Eighe), stones are small and mobility is indicated by bare scree creeping downhill, with

rafts of heather. Elsewhere the Torridonian Sandstone debris is much coarser and often includes large boulders, but numerous breaks in the vegetation cover indicate that instability is present there also.

4. Lower slope forms

These are characterised by debris cones and aprons with apparently greater stability than class 3, and an increased soil cover. For example, more than 1 m of coarse sand was observed in one stream-cut section on the lower slopes of Beinn Dearg. This results in dominance of grass, of which the light photographic appearance contrasts with the much darker tone of Calluna. Stability, however, is not total and there are numerous instances of wash-outs and rotational slipping. On some of the steeper slopes, solifluction terraces are evident.

MORAINIC DEPOSITS

Much of the area is covered by morainic materials, of which five subdivisions are recognised here. Many of these are clearly identifiable by form. Generally, the vegetation consists of Calluna and Erica varieties with grass, Myrica gale, and Eriophorum varieties in the wetter parts. Few good sections exist but observations of isolated examples indicate that the materials are largely angular and unsorted varying in magnitude from boulders of a few metres in diameter to coarse sand. The distinctions which are made here are based on photographic texture.

5. Fine-textured moraine

This consists of small hummocks about 1-3m in height. They are easily distinguished on the photographs as they cause impeded drainage with many bogs and ponds in the hollows. In the characteristic vegetation pattern, the tops of hummocks are virtually bare; heather grows on the sides and the hollows are covered in grass.

6. Coarse-textured moraine

The vegetation pattern is the same as that of fine-textured moraine, but the scale is broader, with large hummocks and ridges in excess of 10 m in height.

7. Lineated moraine

In a few areas the hummocks display definite lineations.

They occur in parallel ridges 2-4 m high, stretching for over 100 m in some cases.

8. Undifferentiated drift

A 'dustbin' class for the substantial smooth areas which are covered by drift but do not show any distinctive form.

9. Boulder fields

The entire Torridon area is littered by boulders, to varying degrees of spatial density. In certain areas, however, boulders are so frequent that the surface below is masked. These areas have been termed boulder fields. They may be of several origins, either ablation moraine or rock-fall debris.

Such a classification implies certain ideas about processes, including a broad distinction between erosion and deposition. Certain categories relate to mass movement, periglacial activity, glacial erosion or glacial deposition. In this particular area, little attention is given to the role of fluvial processes largely because they are not prominent at the scale considered.

In a comparison between the terrain matrix and the altitude matrix these implications are of great importance. But the comparison with the altitude matrix and other matrices derived from altitude can provide some test of the subjectivity of the classification and the accuracy of the mapping.

The classification is based on what appears distinctive to an individual observer, and the mapping has been carried out quite crudely. Some scale distortion due to the considerable altitude variations may remain.

Matrix overlaying

The results of overlaying the terrain matrix on the altitude matrix may now be examined. The frequency distributions of altitude, of gradient and of a measure of local altitude variability are considered.

The first derivative is calculated by a finite difference method as outlined above. Local altitude variability is measured by the modulus of height deviation V_{ij} where

$$4V_{ij} = |z_{ij} - z_{(i+1,j)}| + |z_{ij} - z_{(i-1,j)}| + |z_{ij} - z_{(i,j+1)}| + |z_{ij} - z_{(i,j-1)}|$$

z_{ij} being the altitude at (i,j)

The number in each sample is high; the lowest number is 187. A visual appraisal indicates normality of most of the distributions, the most skewed being 'water' and 'lower slopes'.

With regard to the reliability of the classification it is reassuring that expected relations between the means do materialise. For example, summit terrain has the highest mean altitude and steep scree is higher than lower slope forms. Bare rock appears at all altitudes but morainic terrain types cluster at a fairly low altitude, reflecting their valley-bottom location. The boulder fields of the area occur higher than the morainic deposits. Indeed, their proximity to the mean of steep scree might be taken to suggest a genetic connection.

<u>Statistical characteristics of terrain classes, computed by overlaying the geomorphic matrix on the altitude matrix</u>

			ALTITUDE (m)		GRADIENT	
	class	% area	mean	st.dev.	mean	st.dev.
water	0	3.96	367.9	79.71	14.76°	12.48°
rock	1	22.75	469.2	158.29	14.57°	10.98°
summit	2	3.02	738.6	94.87	13.10°	11.41°
scree	3	12.85	568.9	142.25	17.23°	12.46°
lower slope	4	21.37	416.0	128.73	16.08°	11.98°
fine moraine	5	8.93	322.7	63.46	15.52°	12.69°
coarse moraine	6	7.21	370.8	58.90	9.31°	8.84°
lineated moraine	7	1.87	366.9	47.21	13.52°	11.49°
drift (undifferentiated)	8	14.99	372.31	118.21	12.63°	9.66°
boulder	9	3.05	513.0	141.50	22.31°	12.29°

The table above shows for each class the proportion of the total area which it occupies, together with the mean and standard deviation of altitudes and gradients measured for points within that class. Standard deviation reflects

the total spread of data in each class, and has a slight positive relationship with area ($r = +0.47$), but this is insignificant at the 95% confidence level. As expected, steep scree has a higher average gradient than lower slope forms. Morainic terrain types are varied, reflecting possible difference in form. Fine-textured moraines have the highest gradient and undifferentiated drift has a comparatively low gradient, which confirms a subjective impression of its low relief. The average modulus of height deviation over 100 m gives a better impression of local altitude variability than does gradient. Steep scree is less variable than lower slope forms, reflecting the predominance of smooth scree fans. The roughness of morainic deposits is picked out, particularly for fine-textured moriane. The exception is lineated moraine, which shows least variation of all the classes. It is possible that, as for coarse-textured moraine, the 100 m sampling interval misses most of the variation.

It may be concluded from these initial results that the present classification of terrain in the Torridon area is justified, but certain problems remain. The autocorrelation properties of the data have not yet been measured. The contiguity of the various classes also requires measurement: does a class occur in a large, coherent body or as a number of smaller patches? This becomes important when calculating derivatives, for it may be necessary to use contiguous values outside the areal limits of the class.

If the average size of patch is small, boundary effects may have undue influence. For example, the gradient of over 15° attributed to areas of water is presumably due to this effect. The matrix resolution is crucial, as has been seen; alteration of this may result in variations at different scales being picked up. Increased resolution will not necessarily reduce the boundary effect, for more intricate boundaries may be picked out.

References

BOEHM, B.W., 1967, Tabular representations of multivariate functions - with applications to topographic modelling; Proc., Assoc. Computing Machinery National Meeting, p.403-415

EVANS, I.S., and I. BAIN, 1973, Statistical characterization of altitude matrices by computer, Report 1 on Grant DA-ERO-591-73-G0040, 16pp.

GOODIER, R. and GRIMES, B.H., 1970, The interpretation and mapping of vegetation and other ground surface features from air photographs of mountainous areas in North Wales; Photogrammetric Record, 6 (36), p. 553-566.

JENKINS, G.M., 1961, General considerations in the analysis of spectra; Technometrics 3 (2) p. 133-166, reprinted as p. 98-131 in E. PARZEN (ed.) 'Time Series Papers' Holden-Day, 1967.

RAYNER, J.N., 1971, An introduction to spectral analysis; Pion (London)

RAYNER, J.N., 1972, The application of harmonic and spectral analysis to the study of terrain; p.283-302 in R.J. CHORLEY (Ed.), Spatial Analysis in Geomorphology; Methuen (London).

RAYNER, J.N. and G. McCALDEN, 1972, Computing programs for cross-spectral analysis in two dimensions; Ohio State University, Department of Geography, Discussion Paper 27, 66pp.

TOBLER, W., 1970, Selected computer programs from the Department of Geography, University of Michigan, Ann Arbor (U.S.A.), 162pp.

TUKEY, J.W., 1961, Discussion, emphasising the connection between analysis of variance and spectrum analysis; Technometrics 3 (2) p.191-219.

TOBLER, W.R. 1969, An analysis of a digitalized surface; In Davis, C.M. (Ed.) A Study of the Land Type (U.S. Army Research Office, Durham, North Carolina, Contract DA-31-124-ARO-D-456), p.59-76 & 86.